11-2003

# Cryptic Repeated Genomic Recombination During Speciation in Gossypium gossypioides

Richard Clark Cronn
*United States Department of Agriculture*

Randall Lee Small
*University of Tennessee - Knoxville*

Tamara Haselkorn
*Iowa State University*

Jonathan F. Wendel
*Iowa State University*, jfw@iastate.edu

*Evolution*, 57(11), 2003, pp. 2475–2489

# CRYPTIC REPEATED GENOMIC RECOMBINATION DURING SPECIATION IN *GOSSYPIUM GOSSYPIOIDES*

RICHARD CRONN,[1] RANDALL L. SMALL,[2] TAMARA HASELKORN,[3,4] AND JONATHAN F. WENDEL[3,5]
[1]*Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, Oregon 97331*
[2]*Department of Botany, University of Tennessee, Knoxville, Tennessee 37996*
[3]*Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa 50011*

*Abstract.*—The Mexican cotton *Gossypium gossypioides* is a perplexing entity, with conflicting morphological, cytogenetic, and molecular evidence of its phylogenetic affinity to other American cottons. We reevaluated the evolutionary history of this enigmatic species using 16.4 kb of DNA sequence. Phylogenetic analyses show that chloroplast DNA (7.3 kb), nuclear ribosomal internal transcribed spacers (ITS; 0.69 kb), and unique nuclear genes (8.4 kb) yield conflicting resolutions for *G. gossypioides*. Eight low-copy nuclear genes provide a nearly unanimous resolution of *G. gossypioides* as the basalmost American diploid cotton, whereas cpDNA sequences resolve *G. gossypioides* deeply nested within the American diploid clade sister to Peruvian *G. raimondii*, and ITS places *G. gossypioides* in an African (rather than an American) clade. These data, in conjunction with previous evidence from the repetitive fraction of the genome, implicate a complex history for *G. gossypioides* possibly involving temporally separated introgression events from genetically divergent cottons that are presently restricted to different hemispheres. Based on repetitive nuclear DNA, it appears that *G. gossypioides* experienced nuclear introgression from an African species shortly after divergence from the remainder of the American assemblage. More recently, hybridization with a Mexican species may have resulted in cpDNA introgression, and possibly a second round of cryptic nuclear introgression. *Gossypium gossypioides* provides a striking example of the previously unsuspected chimeric nature of some plant genomes and the resulting phylogenetic complexity produced by multiple historical reticulation events.

*Key words.*—*Gossypium gossypioides*, hybridization, introgression, phylogenetic incongruence, speciation.

Received November 27, 2002. Accepted June 3, 2003.

Interspecific hybridization is presumed to be a common occurrence in flowering plants, with a recent estimate of 27,500 hybrids among the approximately 250,000 described plant species worldwide (Arnold 1979; Rieseberg 1997). Historically, the detection of hybrids and their derivatives relied upon morphological analyses, but modern inferences of hybridization increasingly utilize patterns of incongruence between molecular markers from the cytoplasm (e.g., chloroplast [cp] DNA) and the nucleus (e.g., ribosomal DNA). Among the noteworthy realizations to emerge from comparative molecular phylogenetic analyses is that interspecific hybridization appears to be more prominent in plants than indicated by morphology, as apparently ''nonhybrid'' species commonly possess cytoplasmic markers that are otherwise restricted to genetically divergent species (Arnold 1979; Rieseberg 1997; Wendel and Doyle 1998; Raymond et al. 2002). These data, in conjunction with empirical evidence from model hybrid systems (Arnold 1979; Rieseberg 1997), suggest that interspecific hybridization may play an even more important role in plant speciation than previously recognized.

A striking example of nuclear-cytoplasmic incongruence exists for *Gossypium gossypioides*, a diploid cotton species endemic to Oaxaca, Mexico (Fryxell 1979, 1992; Wendel et al. 1995). This species is one of thirteen diploid New World cottons, a lineage that is genetically (Seelanan et al. 1997; Cronn et al. 2002) and genomically (Endrizzi et al. 1985) differentiated from the remaining 30 plus cotton species native to Africa, Asia, and Australia (Fig. 1). Among New World cottons, eleven species are native to Mexico, with *G. gossypioides* defining the approximate southern limit in the state of Oaxaca (Fryxell 1992). The remaining New World species, *G. raimondii* (native to Peru) and *G. klotzschianum* (from the Galapagos Islands), are isolated from this Mexican center of diversity, most likely due to long-distance (oceanic) dispersal from an ancestral Mexican stock (Fryxell 1971; Wendel and Percival 1990). Due to their chromosomal uniformity, the thirteen New World diploid species (including *G. gossypioides*) are classified in a cytogenetic assemblage called the ''D-genome'' group. Features characteristic of this group include 26 somatic chromosomes and a nuclear 2C DNA content of about 2 pg (Endrizzi et al. 1985). In addition, experimental triploids derived from interploidal crosses between New World species (including *G. gossypioides*) and AD-genome allotetraploids reveal chromosome pairing only with D-genome homoeologues (Menzel and Brown 1954; Endrizzi et al. 1985).

Despite this uncomplicated cartographic and cytogenetic description, molecular evolutionary studies of *G. gossypioides* based on chloroplast DNA (cpDNA; Wendel and Albert 1992), nuclear ribosomal sequences (Wendel et al. 1995; Cronn et al. 1996) and low copy genes (Small and Wendel 2000a; Liu et al. 2001) have yielded three radically different interpretations for the evolutionary history of this species. The first, based upon chloroplast DNA restriction site variation (Wendel and Albert 1992; Fig. 1, circle 1), indicates a sister relationship between *G. gossypioides* and *G. raimondii*, the sole representative of Subsection *Austroamericana*. Despite the geographic separation of these species, this hypothesis is supported by interfertility evidence and morphological data. For example, experimental hybridization of *G. gossypioides* with D-genome cottons shows *G. gossypioides* to be reproductively isolated from all species except *G. raimondii* (Brown and Menzel 1952; Menzel and Brown 1954).

[4] Present address: Department of Zoology, University of Georgia, Athens, Georgia.
[5] Corresponding author. E-mail: jfw@iastate.edu.

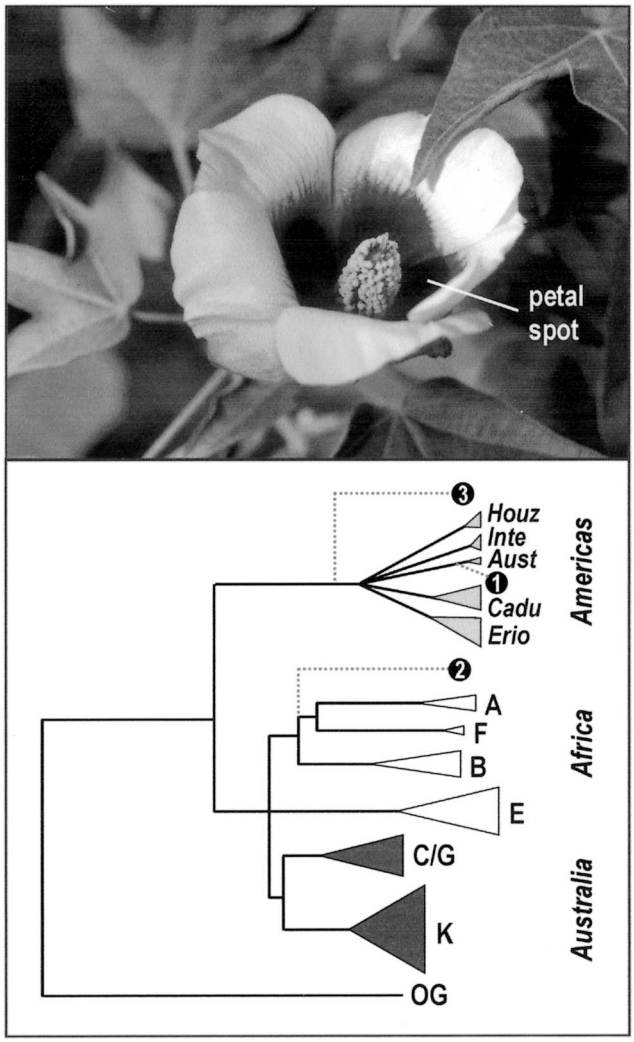



FIG. 1. (Upper panel) Flower from *G. gossypioides*, showing characteristic petal spot pigmentation. (Lower panel) Phylogenetic relationships of genome groups in the cotton genus and the various resolutions of *G. gossypioides*. Molecular data from cpDNA resolves *G. gossypioides* sister to the American Subsection *Austroamericana* (circled ''1''); nuclear ITS resolves *G. gossypioides* sister to African cottons in the A-, F-, and B-genome groups (''2''); 5S rDNA, *AdhA* and *FAD2-1* resolve *G. gossypioides* sister to all New World cottons (''3''). Phylogenetic relationships for the genome groups follow Cronn et al. (2002). Branch lengths within and between genome groups are proportional to the level of sequence divergence exhibited by nuclear ITS (Seelanan et al. 1997), and the height of triangles representing each genome group is proportional to the number of known species from that group. Abbreviations for American D-genome Subsections: *Aust*, *Austroamericana; Cadu*, *Caducibracteolata; Erio*, *Erioxylum; Inte*, *Integrifolia; Houz*, *Houzingenia.*

In addition, these two species also possess a unique petal mutation called ''reverse petal spot,'' where pigment is present on both adaxial and abaxial petal surfaces, as opposed to only the abaxial surface (Fig. 1). This trait is unique among diploid cottons worldwide, and it may represent a synapomorphy uniting these species.

In contrast to the foregoing evidence, ribosomal internal transcribed spacer sequences (ITS) provided an unexpected resolution for *G. gossypioides* (Wendel et al. 1995; Fig. 1, circle 2). Rather than resolving sister to *G. raimondii*, *G. gossypioides* was highly divergent from *all* D-genome cottons, resolving within a clade composed entirely of African species from the A-genome (two extant species), F-genome (one species), and B-genome (three species) groups. Inspection of nucleotide distributions within ITS led to the conclusion that sequences from *G. gossypioides* were chimeric products of a New World D-genome and an African A-genome sequence (Wendel et al. 1995), perhaps arising from an introgression event involving a diploid A-genome species or the A-genome of a New World (AD-genome) allotetraploid cotton. Although this ''African introgression'' hypothesis is complicated by the present-day distribution of A-genome cottons and reproductive barriers presented by interploidal hybridization of diploid *G. gossypioides* and AD-genome allotetraploid cottons (Menzel and Brown 1954), this hypothesis has gained independent support from a recent genus-wide evaluation of multiple repetitive DNA families (Zhao et al. 1998). Among New World cottons, *G. gossypioides* is unique in that it harbors approximately 400,000 copies of ''African enriched'' repetitive DNAs (e.g., abundant in diploid cottons from the A-, F-, and B-genomes) that are absent from the remaining New World and Australian C-, G-, and K-genome cottons.

Finally, the phylogenetic resolution of *G. gossypioides* based on nucleotide variation at the 5S rDNA array (Cronn et al. 1996; Fig. 1, circle 3) was inconsistent with both the ITS *and* the cpDNA data sets, because *G. gossypioides* resolved within the American cotton clade but in a basal phylogenetic position (rather than to sister to *G. raimondii*). This conclusion was initially dismissed as a potential artifact of African introgression (Cronn et al. 1996). Nevertheless, subsequent studies of American cottons using unlinked, low-copy nuclear genes *AdhA* (Small and Wendel 2000a) and *FAD-2-1* (Liu et al. 2001) resolved *G. gossypioides* identically as 5S rDNA, suggesting that this phylogenetic resolution is unlikely to be entirely due to introgressive hybridization.

The conflicting resolutions of *G. gossypioides* provided by molecular data prompt obvious questions. First, what phylogenetic placement best reflects the organismal history of *G. gossypioides*? Is it the sister taxon to Peruvian *G. raimondii*? Does it represent the sister lineage to all New World cottons? What is its relationship to African cottons? The answer to these questions is significant with regard to character evolution in New World cottons, and it may have implications for inferences of long-distance dispersal events. Finally, what events conspired to produce the conflicting phylogenetic signal harbored within cpDNA, rDNA/ITS, and low-copy DNA?

To address these questions, we conducted an extensive evaluation of chloroplast and nuclear sequences from *G. gossypioides* and representative American cottons in order to evaluate the strength of the evolutionary signal in these partitions. Hypotheses initially derived from cpDNA restriction site analysis (Wendel and Albert 1992) were tested by examining 7.3 kb of sequence information from four chloroplast regions (*ndhF*, *matK*, *rpl16* intron, *trnT-trnL* spacer). To evaluate the phylogenetic signal from the nuclear genome, we sampled six low-copy nuclear genes (*CesA1*, *AdhC*, *CesA1b*, *A1341*, *G1121*, and *G1262*; 5.5 kb) to obtain multiple inde-

TABLE 1. Cotton species included in this study.

| Species | Genome | Range |
|---|---|---|
| *G. davidsonii* Kellogg | D | Baja California, Mexico |
| *G. gossypioides* (Ulbrich) Standley | D | Oaxaca, Mexico |
| *G. klotzschianum* Andersson | D | Galapagos Islands |
| *G. raimondii* Ulbrich | D | Peru |
| *G. schwendimanii* Fryxell and S. Koch | D | Michoacán, Mexico |
| *G. trilobum* (DC.) Skovsted | D | Western Mexico |
| *G. turneri* Fryxell | D | Sonora, Mexico |
| *G. hirsutum* L. race Palmeri | AD | Mexico |
| *G. herbaceum* L. | A | South Africa |
| *G. longicalyx* J. B. Hutchinson and Lee | F | East Africa |
| *G. somalense* (Gürke) J. B. Hutchinson | E | East Africa |
| *G. stocksii* Masters | E | East Africa |
| *Gossypioides kirkii* (Mast.) J. B. Hutchinson | — | East Africa/Madagascar |

pendent estimates of evolutionary history. By integrating these data with those derived from ITS (Wendel et al. 1995), 5S rDNA (Cronn et al. 1996), *AdhA* (Small and Wendel 2000a), and *FAD-2-1* (Liu et al. 2001), this evaluation of the affinity of *G. gossypioides* to New World cottons includes about 16.4 kb of sequence information per taxon. Results from this survey show that cpDNA, nuclear ITS, and low-copy nuclear DNA continue to exhibit significantly different patterns of sequence evolution, reflecting their different evolutionary trajectories. After considering the processes that can give rise to pattern incongruence, we conclude that the genomic chimerism of *G. gossypioides* has arisen via hybridization from two temporally divergent (yet mechanistically mysterious) introgression events. The first hybridization event resulted in nuclear introgression of genomically widespread repetitive elements (including the ITS sequence of the 45SrDNA array), whereas the latter led to cytoplasmic introgression. Remarkably, both episodes of introgression involve source populations that are presently located in different hemispheres.

## MATERIALS AND METHODS

### *Plant Materials*

Species examined in this study, including genome group assignments (Endrizzi et al. 1985) and geographic distributions, are listed in Table 1. Accessions selected for analysis are identical to those used in earlier studies (Wendel et al. 1995; Cronn et al. 2002). The cotton genus *Gossypium* L. (Malvaceae) includes about 45 species (Fryxell 1979, 1992) that are differentiated cytogenetically into eight ''genome groups'' (designated ''A'' through ''G,'' and ''K'') that differ in DNA content and chromosome size (Endrizzi et al. 1985). These genome groups are congruent with the recent taxonomic treatments (Fryxell 1979, 1992) and molecular evidence (Wendel and Albert 1992; Seelanan et al. 1997, 1999; Liu et al. 2001; Wendel and Cronn 2003); interrelationships between genome groups have been established using four chloroplast and twelve nuclear genes (Cronn et al. 2002; shown in Fig. 1). The D-genome group includes 13 species that are restricted to the New World. The remaining species (all from the Old World) include 11 cottons native to Africa/Arabia and 16 cottons native to Australia. African/Arabian cytogenetic groups include the A-genome (two species), B-genome (three species), E-genome (five to six species) and F-genome (one species); Australian cytogenetic groups include the C-genome (four species), G-genome (one species), and the K-genome (11 species).

Considerable accumulated evidence supports the monophyly of the six subsections of American cottons (Fryxell 1971; Wendel and Percival 1990; Wendel and Albert 1992; Cronn et al. 1996; Seelanan et al. 1997; Small and Wendel 2000a). Genetic divergence between species in different subsections is high relative to the divergence within species of a subsection (Wendel and Percival 1990; Cronn et al. 1996; Seelanan et al. 1997; Small and Wendel 2000a). For these reasons, we chose one species per subsection as a representative of each major New World lineage. Exemplars include *G. gossypioides* (Subsect. *Selera*; one extant species), *G. raimondii* (Subsect. *Austroamericana*; one species), *G. trilobum* (Subsect. *Houzingenia*; two species), *G. schwendimanii* (Subsect. *Erioxylum*; four species), *G. turneri* (Subsect. *Caducibracteolata*; three species), and *G. davidsonii* or *G. klotzschianum* (Subsect. *Integrifolia*; two species). Orthologous sequences from the D-genome of allotetraploid *G. hirsutum* were also included to evaluate competing hypotheses regarding the most appropriate model D-genome progenitor of allotetraploid cottons (Wendel et al. 1995; Zhao et al. 1998). African cottons were included to represent the sister lineage of American cottons (Cronn et al. 2002; Wendel and Cronn 2003). These include the A-genome species *G. herbaceum*, F-genome *G. longicalyx*, and E-genome species *G. somalense* or *G. stocksii*. Molecular evidence indicates that the genera *Gossypioides* and *Kokia* comprise the sister lineage to *Gossypium* (Seelanan et al. 1997; Wendel et al. 2002), thus *Gossypioides kirkii* was chosen as the outgroup.

### *Chloroplast and Nuclear Loci*

Four chloroplast loci were chosen, including partial *ndhF* gene sequences, the *matK* gene and flanking *trnK* intron, the *rpl16* intron, and the *trnT-trnL* spacer. Chloroplast sequences were amplified, isolated, and sequenced using previously detailed methods (Small et al. 1998; Cronn et al. 2002). Phylogenetic signal from the nuclear genome was evaluated using ten nuclear loci, including six new loci and four previously reported genes. New sequences included three anonymous loci that correspond to *PstI* mapping probes *A1341* (0.7 kb),

TABLE 2. GenBank accession numbers for all taxon/locus combinations.

| Taxon | *matK* | *ndhF* | *rpl16* | *trnT-L* | *A1341* | *AdhA* | *AdhC* |
|---|---|---|---|---|---|---|---|
| *G. gossypioides* | AF520727 | AF520732 | AF520717 | AF520722 | AF520736 | AF182117 | AY125058 |
| *G. schwendimanii* | AF520729 | AF520734 | AF520719 | AF520724 | AF520738 | AF182137 | AY125060 |
| *G. davidsonii* | AF520728 | AF520733 | AF520718 | AF520723 | AF520737 | AF182131 | AY125059 |
| *G. klotzschianum* | — | — | — | — | — | — | — |
| *G. raimondii* | AF403559 | U55335 | AF403101 | AF403549 | AF136815 | AF136459 | AF036568 |
| *G. trilobum* | AF520730 | AF520735 | AF520720 | AF520725 | AF520739 | AF182127 | AY125061 |
| *G. turneri* | AF520731 | U55336 | AF520721 | AF520726 | AF520740 | AF182119 | AY125062 |
| *G. hirsutum (D)* | — | — | — | — | AF136816 | AF090147 | AF036569 |
| *G. hirsutum (A)* | — | — | — | — | — | — | AF036575 |
| *G. herbaceum* | AF403556 | — | — | — | AF136813 | AF136458 | — |
| *G. arboreum* | — | U55331 | AF031451 | AF031433 | — | — | — |
| *G. longicalyx* | AF403561 | U55338 | AF403103 | AF403551 | AF403076 | AF419963 | AF419967 |
| *G. somalense* | AF403560 | — | AF403102 | AF403550 | AF403075 | AF419962 | — |
| *G. stocksii* | — | U55337 | — | — | — | — | — |
| *Gossypioides kirkii* | AF403563 | U55329 | AF403104 | AF403443 | AF201877 | AF201888 | AF169254 |

*G1121* (0.73 kb), and *G1262* (0.89 kb) (Cronn and Wendel 1998; Cronn et al. 1999), partial alcohol dehydrogenase C sequences (*AdhC* = 0.94 kb) (Small et al. 1998; Small and Wendel 2000b), and two partial cellulose synthase genes (*CesA1* = 1.09 kb; *CesA1b* = 1.18 kb). Amplification primers, cycling conditions, and sequencing methods follow previously detailed protocols (Small et al. 1998; Small and Wendel 2000b; Liu et al. 2001; Cronn et al. 2002). As previously noted, *AdhC* exists as a pseudogene in many cotton species (Small et al. 1998; Small and Wendel 2000b), and we were unable to sample this locus from either *G. herbaceum* (A-genome) or the E-genome exemplars. For an A-genome representative, we included the A-genome sequence that has been isolated from the AD-genome allotetraploid species *G. hirsutum* (Small and Wendel 2000b).

In addition to these new sequences, we also reevaluated published data from ITS (0.69 kb; Wendel et al. 1995), consensus sequences of 5S rDNA spacers (0.19 kb; Cronn et al. 1996), alcohol dehydrogenase A gene (*AdhA* = 0.95 kb) (Small and Wendel 2000a,b), and the 5′ UTR intron from a microsomal fatty acid desaturase gene (*FAD2-1* = 1.78 kb; Liu et al. 2001). The criterion used for selecting nuclear loci was that they are known to be low-copy (excepting rDNA) in A- and D-genome diploid cottons (Cronn and Wendel 1998; Brubaker et al. 1999; Small and Wendel 2000b). Additionally, orthological relationships for all loci (excepting *FAD2-1*) have been established by comparative genetic mapping on A-genome and D-genome linkage maps (Cronn and Wendel 1998; Brubaker et al. 1999; Small and Wendel 2000b). GenBank accession numbers for all taxon-locus combinations are summarized in Table 2.

### *Data Analysis*

Sequence alignments from ClustalV (Thompson et al. 1994) was straightforward for most genes, with only three genes displaying complex indel patterns. These regions (excluded from all analyses) include bases 457–492 of the *rpl16* intron (5072–5107 of the cpDNA dataset), 254–435 of the *trnT-trnL* spacer (5988–6169 of the cpDNA dataset), and bases 886–965 and 1157–1179 of *FAD2-1* (6700–6779 and 6971–6993 of the nDNA data). In addition, the 5S gene (1–119 in the 5S dataset; 8298–8417 in the nDNA dataset) was excluded from analysis due to its considerable polymorphism (Cronn et al. 1996). Remaining alignment gaps were treated as missing data.

Phylogenetic analysis of individual and combined datasets utilized maximum parsimony (MP) and maximum likelihood (ML), as implemented by PAUP*4.0b10 (Swofford 1998). Most-parsimonious trees were found using exhaustive searches with equal character weighting. To evaluate support of tree topologies derived from individual and combined datasets, bootstrap resampling (bs) was performed using branch and bound searches with 1000 replicates. Decay analysis (Bremer 1988) was performed by searching exhaustively for all trees up to 20 steps longer than the most-parsimonious, and noting how many steps longer than the shortest tree each clade survived.

For ML analysis, parameter settings for models of sequence evolution were identified using MODELTEST vers. 3.06 (Posada and Crandall 1998). These parameters were estimated for the ITS dataset, combined nDNA (= nuclear datasets minus ITS), and the combined cpDNA dataset. For ITS, the preferred model under the Akaike Information Criterion (AIC) of MODELTEST was a three parameter model with an asymmetric nucleotide rate matrix ($R_{AG}$ = 4.6906; $R_{CT}$ = 7.4671; $R$ = 1.0000 for all other classes) and the proportion of invariable sites set at 0.6054. For nDNA, the preferred model under the AIC was the HKY+I+G (Hasegawa et al. 1985), with an estimated transition/transversion ratio of 1.5186, the proportion of invariable sites set at 0.4264, and a gamma distribution shape parameter of 0.8719. For cpDNA, the preferred model under the AIC was a three parameter model with an asymmetric nucleotide rate matrix ($R_{AC}$ = 1.5209; $R_{AG}$ = 1.1174; $R_{AT}$ = 0.4033; $R_{CG}$ = 0.7926; $R_{CT}$ = 1.1174; $R_{GT}$ = 1.0000), and the proportion of invariable sites set at 0.8091. Estimates of node support under ML were calculated in PAUP* using quartet puzzling (Strimmer and von Haesler 1996; Cao et al. 1998).

To assess congruence among the numerous loci included in this study (Bull et al. 1993), we used the incongruence length difference test (ILD; Farris et al. 1995) to evaluate congruence among pairwise comparisons of individual or combined datasets. This test (implemented in PAUP*4.0 using parsimony informative characters) used 1000 branch and

TABLE 2. Extended.

| *CesA1* | *CesA1b* | *FAD2-1* | *G1121* | *G1262* | 5SrDNA | ITS |
|---|---|---|---|---|---|---|
| AY125065 | AY125070 | AJ244912 | AF520471 | AY125053 | U32032–U32035 | U12724 |
| AY125067 | AY125072 | AY125052 | AF520473 | AY125055 | U32036–U32039 | U12734 |
| AY125066 | AY125071 | — | AF520472 | AY125054 | U32054–U32055 | U12729 |
| — | — | AJ244910 | — | — | — | — |
| AF139444 | AF139449 | AJ244913 | AF139434 | AF061089 | U32074–U32077 | U12718 |
| AY125068 | AY125073 | AJ244909 | AF520744 | AY125056 | U32056–U32058 | U12723 |
| AY125069 | AY125074 | AJ244911 | AF520745 | AY125057 | U32066–U32067 | U12726 |
| AF139445 | AF139450 | AJ244923 | AF139435 | AF061087 | U32085, U39499 | U12719 |
| — | — | — | — | — | — | — |
| AF139442 | AF139447 | AJ244915 | AF139432 | AF061085 | U32006–U32010 | U12713 |
| — | — | — | — | — | — | — |
| AF419972 | AF419976 | AF403072 | AF377874 | AF402306 | — | U12722 |
| AF419971 | AF419975 | AJ244916 | AF377873 | AF402305 | — | U56809 |
| — | — | — | — | — | AY125064 | — |
| AF201886 | AF201887 | AF403073 | Af201884 | AF201885 | — | U56783 |

bound searches in which random partitions of equal size were created by sampling sites without replacement from the original data. It is important to note that this test is susceptible to rejecting the null hypothesis of congruence when data partitions show different level of homoplasy (Dolphin et al. 2000; Barker and Lutzoni 2002). Accordingly, spurious rejection of partition homogeneity may be an issue in some of these comparisons. Topological conflict between different phylogenetic hypotheses were evaluated using the Templeton test (for MP-based analyses; Templeton 1983) and the SH test (for ML-based analyses; Shimodaira and Hasegawa 1999). These tests were performed with PAUP*4.0 using pruned (reduced in taxa) combined datasets to test the alternative topologies obtained from cpDNA, nDNA, and ITS.

## RESULTS

### *Phylogenetic Resolution of Gossypium gossypioides from Chloroplast Loci*

Analysis of the four cpDNA regions yielded 7316 aligned positions, approximately half of which were exonic (3574 bp) and the remainder from introns and intergenic spacers (3742 bp). Summary statistics for all sequences are shown in Table 3. A small proportion of characters were variable (2.6%–7.2% per locus), with 0.55% to 1.5% of characters being phylogenetically informative. Note that allotetraploid *G. hirsutum* was excluded from all cpDNA studies because it contains an A-genome cytoplasm (summarized in Wendel and Albert 1992), and is therefore uninformative with regard to D-genome sequence evolution.

Maximum parsimony analysis of individual chloroplast loci (Fig. 2) provides varying resolution for American cotton species, yielding two (*matK*, *rpl16* intron, *trnT-trnL* spacer) to 12 (*ndhF*) most-parsimonious trees. Because chloroplast loci are linked on a nonrecombinant chromosome, we combined these sequences into a single cpDNA data set for MP and ML analyses. Maximum parsimony analysis of cpDNA yields a single fully resolved tree (Fig. 3) that mirrors the topologies obtained from constituent loci. In this tree, *G. davidsonii* resolves as the sister lineage to all American cottons (matching *matK* and *trnT-trnL*), and *G. gossypioides*, *G. raimondii*, *G. schwendimanii*, and *G. trilobum* are resolved as a monophyletic group (observed in all individual loci). Of greatest relevance to this study, *G. gossypioides* is nested within the core New World clade, resolving weakly as sister to *G. raimondii* (bs = 66%, $d = 0$), but solidly within a clade composed of *G. gossypioides* + *G. raimondii* + *G. schwendimanii* (bs = 91%; $d = +3$). The topology derived from ML analysis of the cpDNA data set (Fig. 3) is identical to the MP tree, and the support for the clade of *G. gossypioides* + *G. raimondii* + *G. schwendimanii* is high (quartet puzzling value = 99%). This result is significant, because it corroborates earlier evidence derived from cpDNA restriction sites (Wendel and Albert 1992) and crossing studies (Brown and Menzel 1952; Menzel and Brown 1954), both of which point to a sister relationship between *G. gossypioides* and *G. raimondii*.

### *Phylogenetic Resolution of Gossypium gossypioides from Nuclear Genes*

Analysis of the ten nuclear regions yielded a total of 9142 aligned positions, of which 2692 were from exons (29.4%) and 6450 (70.6%) were from introns and spacers (Table 3). Over 11% of the characters were variable (6.30%–40.74% per locus) and 2.36% of the positions were phylogenetically informative. Analysis of individual nuclear genes (Fig. 2) provides varying resolution for American cottons, yielding between one (*AdhC*, *CesA1b*, *FAD2-1*) and seven (*A1341*) most-parsimonious trees. With the exception of *G1262* and ITS, nuclear loci resolve American cottons as a monophyletic group distinct from African cottons. This degree of resolution is significant, as it represents the minimum necessary to evaluate whether *G. gossypioides* shows a higher affinity to the D-genome lineage (indicated by cpDNA and 5SrDNA) or the African lineage (e.g., ITS). Of the nuclear loci sampled, seven unambiguously identify *G. gossypioides* as the sister taxon to all American cottons (Fig. 2), with support for this resolution ranging from weak ($d \leq 1$ for *AdhA*, *AdhC*, *CesA1b*, *FAD2-1*, *G1121*) to modest ($d \geq 2$ for *A1341*, 5S rDNA). For *G1262*, *G. gossypioides* resolves in a basal trichotomy and is not included within the larger D-genome clade; however, the E-genome exemplar *G. somalense* does resolve within the D-genome clade, apparently creating a paraphyletic D-

TABLE 3. Description of regions sequenced from plastid and nuclear genomes.

| Locus | Aligned length[1] | Exon/Intron length[2] | Alignment gaps | Simple gaps[3] | Complex gaps[3] | % VAR[4] | % PI[4] |
|---|---|---|---|---|---|---|---|
| *ndhF* | 2060 | 2060/0 | 9 | 9 | 0 | 2.6 | 0.63 |
| *matK* | 2552 | 1512/1040 | 146 | 95 | 51 | 2.9 | 0.67 |
| *rpl16* | 1119 | 0/1119 | 59 | 42 | 17 | 3.3 | 0.55 |
| *trnT-trnL* | 1585 | 0/1585 | 457 | 188 | 269 | 7.2 | 1.5 |
| Plastid total | 7316 | 3574/3742 | 671 | 334 | 337 | 3.7 | 0.75 |
| *A1341* | 703 | 0/703 | 23 | 23 | 0 | 7.97 | 1.42 |
| *AdhA* | 953 | 611/342 | 22 | 22 | 0 | 6.92 | 1.26 |
| *AdhC* | 940 | 433/507 | 0 | 0 | 0 | 15.30 | 2.87 |
| *CesA1* | 1087 | 593/494 | 15 | 15 | 0 | 7.45 | 1.20 |
| *CesA1b* | 1176 | 566/610 | 73 | 69 | 4 | 9.44 | 0.77 |
| *FAD2-1* | 1783 | 0/1783 | 797 | 547 | 250 | 15.86 | 4.09 |
| *G1121* | 732 | 325/407 | 37 | 37 | 0 | 6.83 | 1.50 |
| *G1262* | 888 | 888/0 | 0 | 0 | 0 | 6.30 | 0.90 |
| ITS | 691 | 167/524 | 26 | 26 | 0 | 17.22 | 5.06 |
| 5S spacer | 189 | 0/189 | 19 | 8 | 11 | 40.74 | 10.58 |
| Nuclear total | 9142 | 2692/6450 | 1012 | 747 | 265 | 11.35 | 2.36 |

[1] Lengths are in bp.
[2] Intron totals include spacers.
[3] Simple gaps are either autapomorphic or have an identical start and end point; complex gaps show a different start or end point, indicating overlapping insertion or deletion events.
[4] Percent variable (VAR) and phylogenetically informative (PI) sites include outgroup comparisons. Values do not include sites excluded due to alignment gaps or ambiguity.

genome (albeit with no branch support). Comparisons among loci show that resolution of the remaining D-genome species is variable, due in large part to the limited sequence divergence separating these lineages.

Our MP and ML analysis of previously published ITS data with a more appropriate outgroup (*Gossypioides kirkii* rather than *G. robinsonii*, as per Wendel et al. 1995) resolves *G. gossypioides* sister to African *G. herbaceum* and *G. longicalyx* in the same manner as earlier studies (Figs. 2, 3). This association is supported by high bootstrap and decay values (bs = 99%; $d = +7$) in MP analysis and a high quartet puzzling score (100%) in ML analysis.

Although our topological conclusions for ITS match those of Wendel et al. (1995), our interpretation of the ''hybridity'' of this sequence is substantially different. Wendel et al. (1995) observed that ITS sequences from three accessions of *G. gossypioides* were a mosaic of A- and D-genome specific nucleotides. These authors suggested that the pattern arose from multiple recombination events between arrays in an early A × D hybrid ancestor. This finding was challenged by Buckler et al. (1997), who noticed that the ratio of homoplasious characters to total character difference was low, and that ''D-genome specific'' nucleotides could be interpreted as plesiomorphic character states. By polarizing the pattern of sequence change with *Gossypioides kirkii* (Table 4), we find that the sites uniting *G. gossypioides* with New World species are shared ancestral states, as suggested by Buckler et al. (1997). In our comparison, ITS from *G. gossypioides* shows no apparent synapomorphies with *any* American cotton species, whereas it demonstrates seven apparent synapomorphies with African A-, B-, and F-genome species (which comprise a well-supported monophyletic lineage; Cronn et al. 2002). Based on this new insight, we conclude that ITS from *G. gossypioides* is best described as a derived ''African'' ITS sequence that shows no evidence of interaction with New World D-genome rDNA.

Organellar genes are inherited as a linked unit, so introgression and lineage sorting can be expected to affect the entire genome. Thus, different cpDNA sequences should converge on a common phylogenetic hypothesis (e.g., Fig. 2). Nuclear genes, however, evolve in the presence of independent assortment and recombination. For this reason, unlinked nuclear loci contain congruent phylogenetic signal only as a consequence of shared phylogenetic history. Consequently, the unlinked nuclear loci included in this study provide truly independent sources of evidence for phylogenetic analyses.

Despite the independence of these datasets, results of the ILD test (Table 5) showed that only four significant differences were detected in the 45 possible pairwise nuclear gene comparisons, with all significant tests involving ITS. In pairwise comparisons with ITS, data sets from *AdhA*, *AdhC*, *CesA1*, and *FAD2-1* showed conflict ($P \leq 0.05$ for *AdhA*, *AdhC*, *CesA1*, and *FAD2-1*); the remaining nonsignificant comparisons to ITS included three of the least informative genes (*Ces1b*, *G1121*, *G1262*). Although the ILD test has been recently demonstrated to be an unreliable guide of dataset incongruence (Dolphin et al. 2000; Barker and Lutzoni 2002), the repeated pattern of conflict between ITS and the other nuclear genes highlights the uniqueness of the ITS resolution. Because the ITS resolution of *G. gossypioides* appeared at odds with four nuclear loci, and the remaining loci showed minimal conflict inter se (Table 5), we chose to pool all nuclear sequences except ITS to evaluate the resolution of *G. gossypioides* derived from nuclear data (nDNA). When this combined nDNA dataset is compared to the ITS dataset using the ILD test (using only parsimonious informative characters as described in Materials and Methods), these datasets exhibit significant pattern incongruence ($P = 0.001$). However, if *G. gossypioides* is excluded from this comparison, the ITS and nDNA datasets show no detectable conflict ($P = 1.000$). Critically, removal of any single D-genome species besides *G. gossypioides* makes no change in the observed

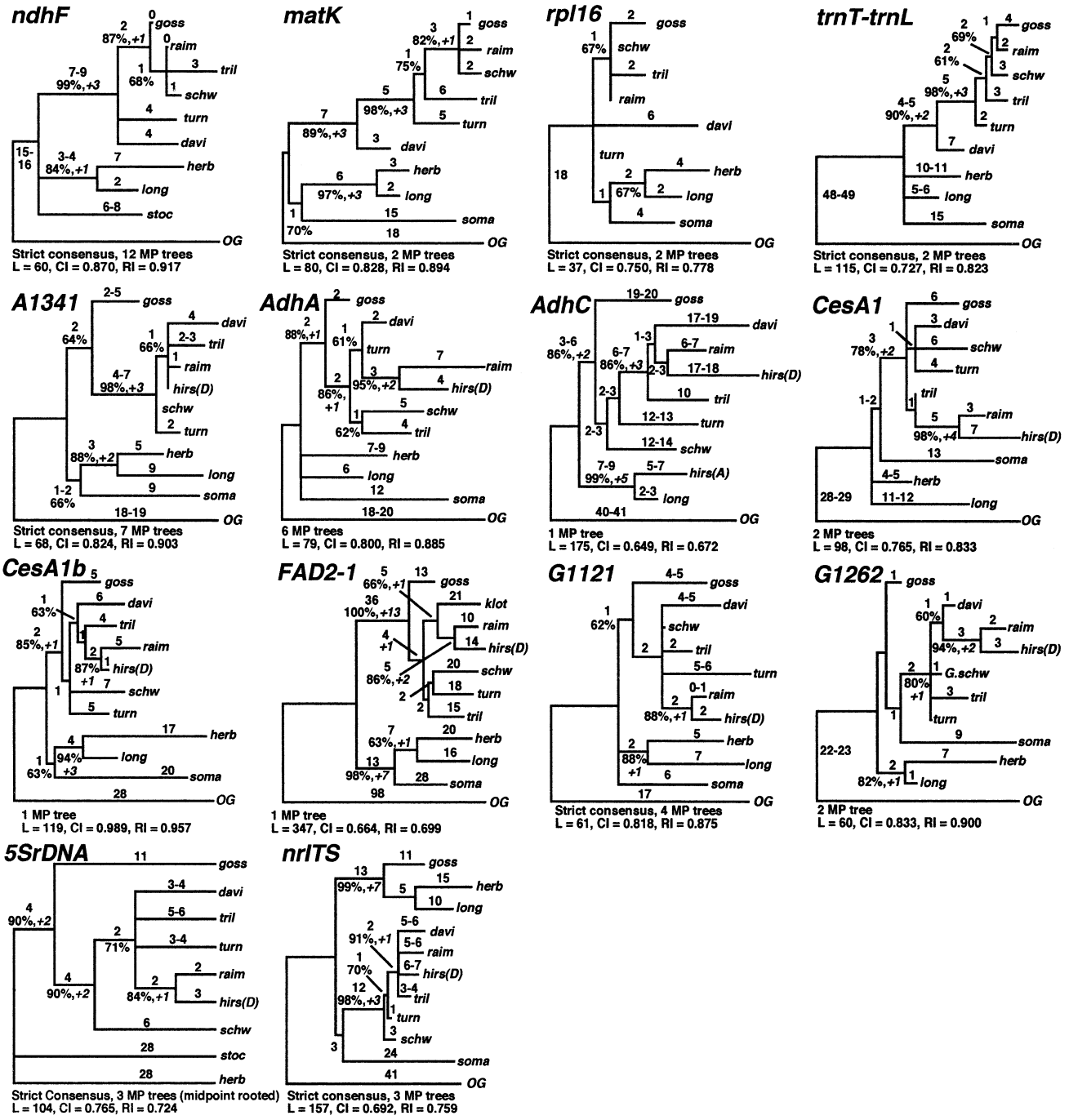



FIG. 2. Maximum-parsimony estimates of phylogenetic relationships of New World cottons from individual chloroplast (upper row) and nuclear genes. Values listed on MP trees include the inferred branch length, bootstrap percentages greater than 60%, and decay indices (in italics). Tree length (L); rescaled consistency index (CI, autapomorphies excluded); and retention index (RI) are provided below trees. Taxon names are abbreviated with the first four letters of the species name, with the genome of tetraploid *G. hirsutum* indicated in parentheses. The outgroup is designated as ''OG.''

level of dataset incongruence ($P = 0.001$ for all other comparisons). Clearly, the primary source of conflict between ITS and other nuclear loci is the unusual resolution of *G. gossypioides* alone.

Maximum parsimony analysis of the combined nDNA dataset resulted in two MP trees (Fig. 3) that were similar to trees derived from constituent loci. With nDNA, *G. gossypioides* is clearly resolved as the sister lineage to all extant New World cottons in both MP and ML analyses (Fig. 3), contradicting the resolution of this species from cpDNA and ITS. Node support for this *G. gossypioides* resolution is high in both MP (bs = 100%, $d = >20$) and ML (puzzle scores

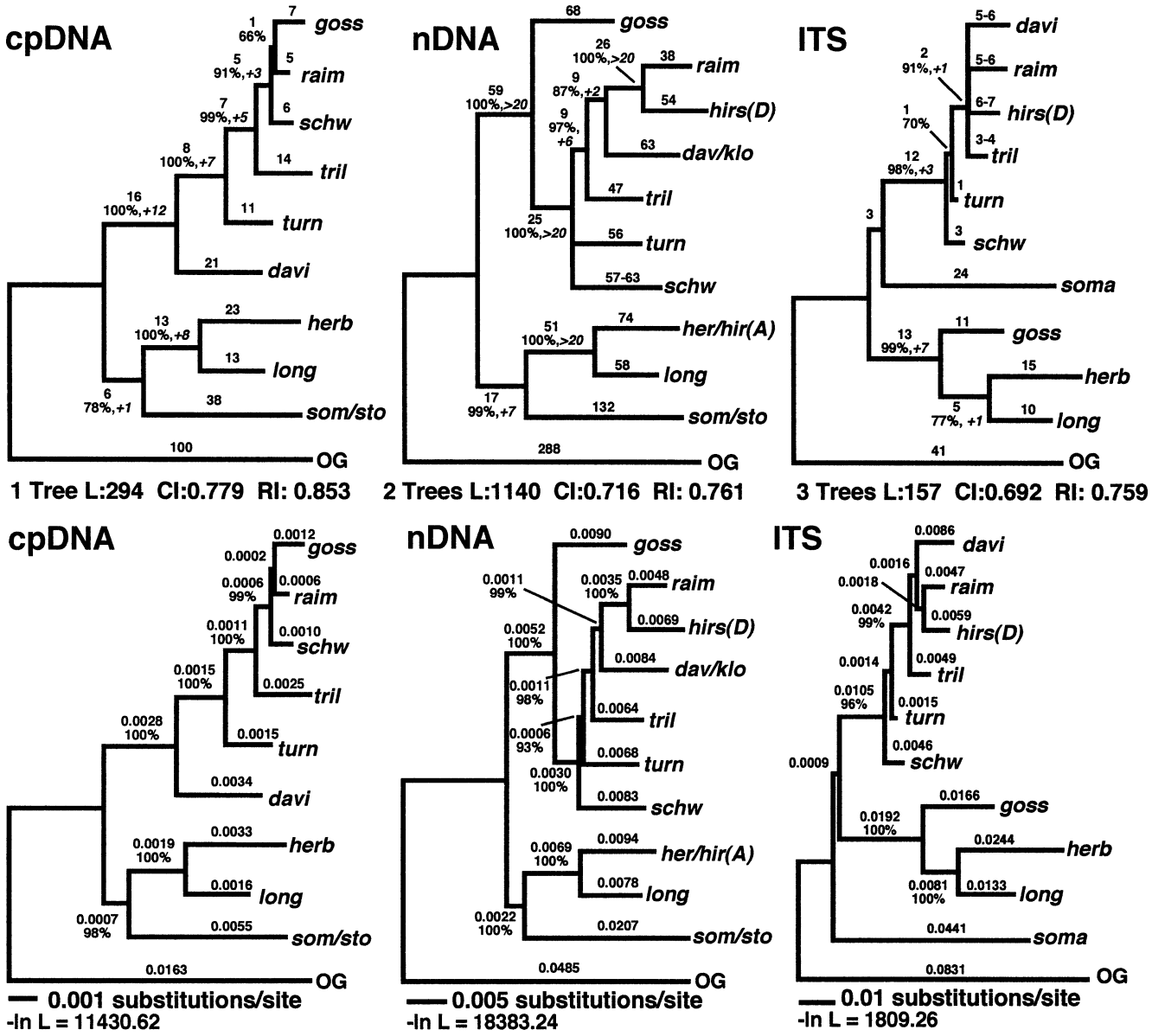



FIG. 3. Maximum-parsimony (upper row) and maximum-likelihood (lower row) estimates of phylogenetic relationships from combined chloroplast (cpDNA) genes, combined nuclear (nDNA) genes excluding ITS, and ITS alone. Values listed for MP trees include the inferred branch length, bootstrap percentages greater than 60%, and decay indices (in italics). Tree length (L), rescaled consistency index (CI), and retention index (RI) are provided below trees. Values listed for ML trees include branch lengths (as expected number of changes per site) and quartet puzzling values for node support. Tree scores (−ln likelihood) are shown under each tree. Taxon names are either abbreviated with the first four letters of the species name or the first three letters when two related taxa represent the lineage. The outgroup is designated as ''OG.''

of 100%) analyses. Another important distinction between cpDNA and nDNA (including ITS) involves the resolution of *G. davidsonii* (and apparently *G. kloztschianum*). This species resolves sister to all New World cottons in cpDNA analyses, with robust support from MP (bs = 100%; $d = +7$) and ML (puzzle score = 100%). In contrast, nDNA resolves *G. davidsonii*/*G. klotzschianum* sister to *G. raimondii* + *G. hirsutum* with modest support (bs = 87%; $d = +2$). In this regard, ITS provides a similar resolution, with *G. davidsonii* resolving sister to *G. raimondii*, *G. hirsutum*, and *G. trilobum* (bs = 91%; $d = +1$). Finally, the D-genome of AD-genome allopolyploid *G. hirsutum* shows the highest affinity to *G. raimondii* based on nDNA (Figs. 2, 3). This result is at least minimally supported by eight of the ten loci examined, with three loci (*AdhA*, *CesA1*, *G1262*) showing bootstrap values over 90%. This demonstrates that the D-genome of *G. hirsutum* and *G. raimondii* share a more recent common ancestor than do those of *G. hirsutum* and *G. gossypioides*, contra-

TABLE 4. All ingroup ITS synapomorphies from select *Gossypium* species. Synapomorphic positions uniting *G. gossypioides* with other cotton species are shown in bold; synapomorphies uniting *G. gossypioides* with African species are in bolded italics.

| Species | Genome group | GenBank number[1] | Aligned position: 23447 77789 90000 01244 45569 12233 46898 88994 89034 66 / 95091 67891 41456 71837 90866 41715 01394 58575 02952 09 |
|---|---|---|---|
| *G. raimondii* | D | U12718 | CACCC TCCGT CC**CC**A GCGGC G**G**CGC CAGGT CGCCA TGACC **C**GTTC GC |
| *G. schwendimanii* | D | U12734 | T**C**CCC T**C**CGT CCT**C**G GCAGT G**G**CGC CGGGT CGCCA TGACC **C**GTTT GC |
| *G. trilobum* | D | U12723 | T**C**CCC T**C**CGT CC**CC**A GCGGC G**G**CGC CAGGG CGCCA TGACC **C**GTTT GC |
| *G. turneri* | D | U12726 | T**C**CCC T**C**CGT CCT**C**G GCAGC G**G**CGC CAGGT CGCCA TGACC **C**GTTT GC |
| *G. davidsonii* | D | U12729 | T**C**CCC T**C**CGT CCT**C**A GCGGC G**G**CGC CRGAC CTCCA TGACC **C**GCTC GC |
| *G. gossypioides* | D | U12724 | T**C**T***TG*** TCC***A***C ***G***CC**C**G TTA***A***T A**G**C***TT*** CA***A***GT C**A**TCC ***C***AG***AT*** **C**GTTT GT |
| *G. herbaceum* | A | U12713 | T**C**T***TG*** TTGCC ***G***CC**C**G GTAGT A**G**T***T***C TA***A***AC C**A**TTC ***C***TG***A***C **C**GTTT AC |
| *G. anomalum* | B | U56806 | T**C**T***TG*** T**C**GC ***G***CC**C**G ATAAT A**G**T***T***C CA***A***GA T**A**TTC ***C***TG***AT*** AGTTT GC |
| *G. longicalyx* | F | U12722 | T**C**T***TG*** T**C**GGC ***G***TC**C**G GTAGT A**G**C***TT*** TA***A***GT T**A**TTC ***C***TG***A***C **C**CTTT GT |
| *G. somalense* | E | U12809 | T**C**TCC A**C**G***A***C CCTTG ACAGT AACGC CGGAT C**A**CCC TGGCT **C**CTTT GC |
| *G. robinsonii* | C | U12710 | C**C**TCC A**C**CGC CT**CC**G GCAGT AACGC CAGGT C**A**CCC TGGCG **C**CTCT AC |
| *G. bickii* | G | U56787 | C**C**TCC G**C**CGC CCT**C**G GCAGT AACGC CAGGT C**A**CCC TGGCC **C**CCCT GC |
| *Gossypioides kirkii* | — | U56783 | TATCC YTCGC CCTTG GTAGT AACGC CAGGT CGTCC TGGCT AGTTT GT |

[1] All sequences were previously described in Wendel et al. (1995) and Seelanan et al. (1997).

dicting earlier inferences derived from ITS and repetitive DNA (Wendel et al. 1995; Zhao et al. 1998; summarized in Wendel and Cronn 2003).

### *Support for Alternative Phylogenetic Resolutions for G. gossypioides*

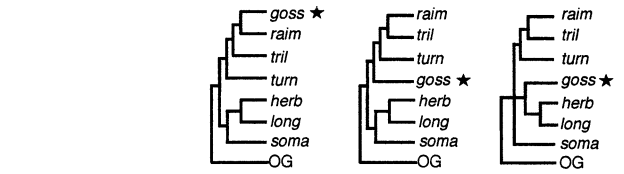
To evaluate support for the competing resolutions of *G. gossypioides* revealed by cpDNA, nDNA, and ITS, we explored statistical support for these topologies using the ML-based SH test (Shimodaira and Hasegawa 1999) and the MP-based Templeton test (Templeton 1983). These tests were performed on datasets that excluded two phylogenetically variable diploid taxa (*G. davidsonii* and *G. schwendimanii*) and tetraploid *G. hirsutum* (which possess an A-genome cytoplasm; see Wendel and Albert 1992). This narrowed taxon sampling allows us to focus our examination on the conflict induced by various resolutions of *G. gossypioides*.

Results from both the ML- and MP-based tests show a high degree of incongruence between the various resolutions of *G. gossypioides* from these three datasets. For example, enforcing the cpDNA resolution (e.g., *G. gossypioides* sister to *G. raimondii*) on the nDNA dataset produces a significant increase in the ML score ($\Delta\ln L = +70.64$, $P = 0.010$; Fig. 4, upper panel). Similarly, enforcing the ITS resolution (e.g., *G. gossypioides* sister to *G. herbaceum*) onto nDNA data yields an increase of even greater significance ($\Delta\ln L = +121.85$, $P \leq 0.001$). Results from the MP-based Templeton test show an identical trend (Fig. 4, lower table), as the nDNA-based resolution appears significantly different from either the cpDNA ($z = -4.7676$; $P \leq 0.001$) or ITS ($z = -6.5320$; $P \leq 0.001$) resolutions. Analysis of topological incongruence using the ITS dataset yields similar results as the nuclear dataset with the ML-based SH test. The pruned ITS dataset yields an ML score of 1721.21; imposing either the cpDNA topology ($\Delta\ln L = +50.00$; $P \leq 0.001$) or nDNA topology ($\Delta\ln L = +23.78$; $P = 0.022$) returns a significant result. These comparisons shows that the nuclear and ITS datasets are sufficiently robust that they statistically reject the alternative hypotheses for *G. gossypioides* shown in Figure 4.

In contrast to these datasets, cpDNA did not appear universally robust with regard to topological change as assessed by the SH test. The pruned cpDNA dataset yields an ML score of 11,202.77 (Fig. 4, upper table). When the ITS resolution of *G. gossypioides* is imposed on this dataset, the resulting $\Delta\ln L$ is $+160.46$, which is significant at $P \leq 0.001$. In this regard, the ITS resolution of *G. gossypioides* is sufficiently different from the cpDNA data that the two are clearly in conflict. However, when the nDNA resolution of *G. gossypioides* is imposed upon the cpDNA dataset, the $\Delta\ln L$

TABLE 5. Incongruence length difference (ILD) test results from all pairwise combinations of cpDNA and nDNA data. *P*-values are listed and comparisons returning a $P \leq 0.05$ are shown in bold.

| Gene | *A1341* | *AdhA* | *AdhC* | *CesA1* | *CesA1b* | *FAD2* | *G1121* | *G1262* | 5SrDNA |
|---|---|---|---|---|---|---|---|---|---|
| *A1341* | — | | | | | | | | |
| *AdhA* | 0.29 | — | | | | | | | |
| *AdhC* | 1.00 | 0.10 | — | | | | | | |
| *CesA1* | 0.33 | 0.12 | 0.45 | — | | | | | |
| *CesA1b* | 1.00 | 0.15 | 0.84 | 0.66 | — | | | | |
| *FAD2* | 0.83 | 0.63 | 0.53 | 0.61 | 0.45 | — | | | |
| *G1121* | 1.00 | 1.00 | 1.00 | 0.34 | 1.00 | 1.00 | — | | |
| *G1262* | 0.60 | 1.00 | 1.00 | 1.00 | 0.07 | 0.64 | 0.59 | — | |
| 5SrDNA | 1.00 | 0.20 | 0.39 | 0.57 | 1.00 | 0.30 | 1.00 | 0.39 | — |
| ITS | 0.57 | **0.01** | **0.05** | **0.02** | 0.08 | **0.01** | 0.93 | 1.00 | 0.36 |




| Data Set | SH Result | cpDNA Resolution | nDNA Resolution | ITS Resolution |
|---|---|---|---|---|
| cpDNA data | - ln L | 11202.77 | 11234.51 | 11363.23 |
| | Δ ln L | *(best)* | 31.73 | 160.46 |
| | *P* | - | 0.107 | **<0.001** |
| nDNA data | - ln L | 17182.52 | 17111.88 | 17233.72 |
| | Δ ln L | 70.64 | *(best)* | 121.85 |
| | *P* | **0.010** | - | **<0.001** |
| ITS data | - ln L | 1471.21 | 1744.99 | 1721.21 |
| | Δ ln L | 50.00 | 23.78 | *(best)* |
| | *P* | **<0.001** | **0.022** | - |

| Data Set | WSR Result | cpDNA Resolution | nDNA Resolution | ITS Resolution |
|---|---|---|---|---|
| cpDNA data | L | 265 | 273 | 300 |
| | *z* | *(best)* | -2.3094 | -5.6045 |
| | *P* | - | **0.021** | **<0.001** |
| nDNA data | L | 985 | 956 | 1004 |
| | *z* | -4.7676 | *(best)* | -6.5320 |
| | *P* | **<0.001** | - | **<0.001** |
| ITS data | L | 164 | 154 | 147 |
| | *z* | -3.9001 | -2.333 | *(best)* |
| | *P* | **<0.001** | **0.020** | - |

FIG. 4. Evaluation of alternative topologies for *G. gossypioides* indicated by cpDNA, nDNA (without ITS), and ITS data sets using ML- (upper table) and MP-based incongruence tests (lower table). ML scores (−ln L) for user-input trees were determined for all datasets as described in Materials and Methods. Changes in ML scores (Δln L) were evaluated for significance using the SH likelihood ratio test (Shimodaira and Hasegawa 1999). Increases in MP tree length (L) were evaluated for significance using the test of Templeton (Templeton 1983). *P*-values shown in bold indicate significance at the 0.05 threshold or lower.

is smaller (+31.73) and insignificant ($P = 0.107$) based on the SH test. This finding contrasts results obtained from the MP-based Templeton test (Fig. 4, lower panel) where the same exercise (cpDNA data constrained to the nDNA topology) increases the treelength from 265 to 273 steps and is significant at $P = 0.021$.

In this instance of ''near-incongruence,'' it is important to note that cpDNA records two potentially informative indels for *G. gossypioides*, both of which serve as synapomorphies with *G. raimondii*. One is a 4 bp insertion in the *trnK* intron (positions 575–579 of the *matK* dataset and the cpDNA alignment) that is present in *G. gossypioides*, *G. raimondii*, and *G. schwendimanii*, and absent from all other species. The second is a 5 bp deletion in *trnT-trnL* (positions 117–121 in the *trnT-trnL* alignment, 5850–5854 of the cpDNA alignment) that is restricted to *G. gossypioides* and *G. raimondii*. If these phylogenetically informative indels are scored as binary characters and included in the pruned cpDNA dataset, the change in the tree length becomes more pronounced, with the cpDNA data returning a treelength of 267 steps for the cpDNA topology (from Fig. 4) and 276 steps for the nDNA topology. As expected, adding these two characters increases in the sum of signed ranks (from +65 to +77 for positive ranks; −13 to −14 for negative ranks) and returns a Templeton test result that is significant at $P = 0.013$. Because the SH and Templeton tests return different results, and significance levels are influenced by the inclusion of potentially informative indels (in the MP-based Templeton test), we interpret the cpDNA results as indicative of ''probable phylogenetic conflict'' between the cpDNA and nDNA resolutions.

## DISCUSSION

Recent years have witnessed an increasing awareness of the importance of hybridization in plant evolution, both at the diploid (Arnold 1979; Rieseberg 1997) and polyploid levels (Soltis and Soltis 1999; Wendel 2000; Osborne et al. 2003). Although the reticulate ancestry of many species can be inferred from morphological evidence, in many cases interspecific ancestry remains unsuspected until serendipitously revealed through molecular phylogenetic analysis. In particular, incongruence between cpDNA-based and nuclear (usually ITS)-based gene trees is responsible for the majority of insights in this arena (Arnold 1979; Rieseberg 1997; Wendel and Doyle 1998). In these instances, the most robust inferences of ancient reticulation will emerge from multiple, independent phylogenetic data sets. We have taken this approach to study the evolutionary history of *G. gossypioides*, a narrowly distributed Mexican species long thought to represent a conventional D-genome taxon, but which may posses a complex reticulate ancestry. An important conclusion of this study is that phylogenetic inferences initially derived from cpDNA restriction sites (Wendel and Albert 1992), ITS sequences (Wendel et al. 1995), and nDNA sequences (Cronn et al. 1996; Small and Wendel 2000a; Liu et al. 2001) are identical to those obtained from more extensive sampling of nucleotide sequences (cpDNA, nDNA), reanalysis with more appropriate outgroups (ITS), and different reconstruction methods. Thus, the divergent resolutions are robust, even though radically different interpretations are indicated by each dataset. Thus, these trees are ''correct'' to the extent that they clearly resolve the evolutionary trajectory of *G. gossypioides* relative to other New World cottons; nevertheless, their conflict offers many compelling examples of ''hard'' incongruence (Seelanan et al. 1997).

Our present analysis considers eleven independent sources of data, one comprising four linked-chloroplast genes, and ten others representing unique nuclear loci that reside on a minimum of six of the 13 haploid chromosomes present in diploid cottons (Brubaker et al. 1999). A critical observation is that 5S rDNA and the single-copy nuclear genes converge upon a common resolution for *G. gossypioides*, namely, that it is the sister taxon of all remaining New World cottons. In light of the independence of these estimates, we suggest that this phylogenetic interpretation reflects the pattern of evolution across the nuclear genome, and that it best represents the organismal affinity of *G. gossypioides*. Given this interpretation, how then do we accommodate the different resolutions derived from the cpDNA and ITS datasets, as well as the apparent colonization of repetitive elements into the *G. gossypioides* nuclear genome that are presently restricted to African *Gossypium* (Wendel et al. 1995; Zhao et al. 1998)? In our opinion, introgression arising from two interspecific hybridization events is the most likely source of conflict (Fig. 5), even though such a scenario seems improbable based on

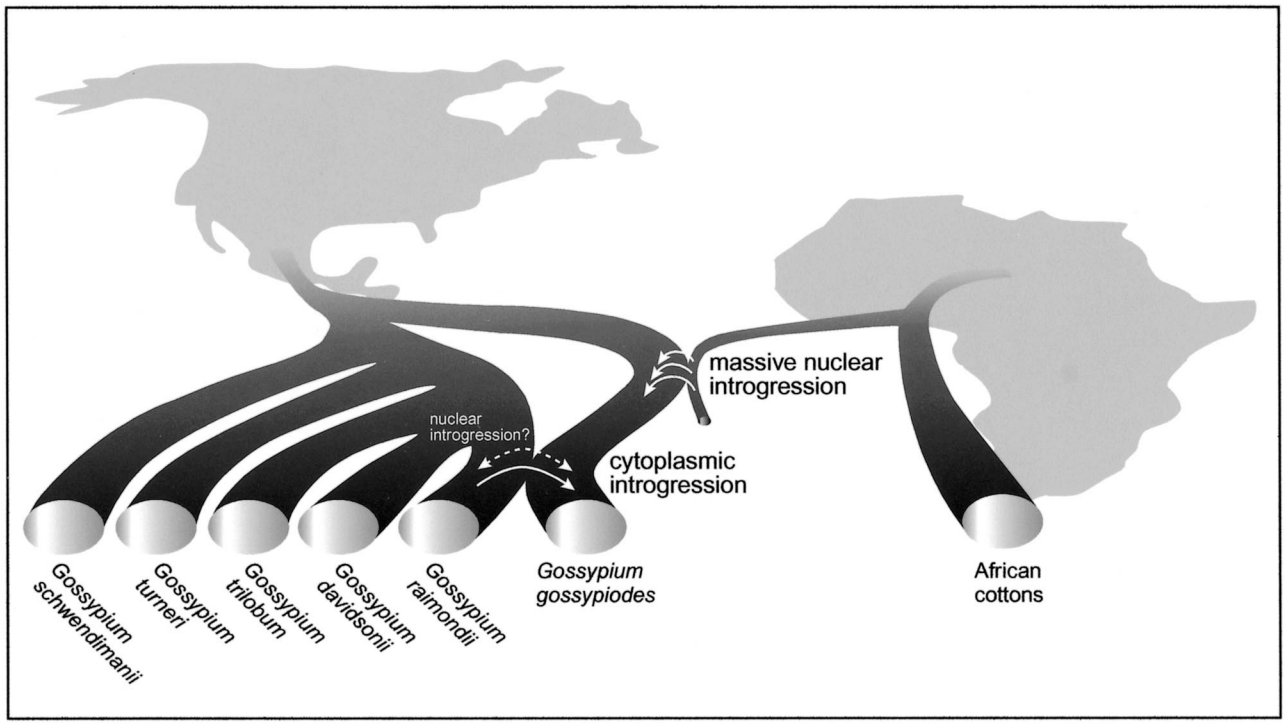



FIG. 5. Phylogenetic relationships among New World cottons and the hypothesized events contributing to the reticulate ancestry of *Gossypium gossypioides*. After the progenitor of *G. gossypioides* diverged from other New World cottons, this entity hybridized with an unknown African species that transiently colonized the New World. This hybridization event resulted in the widespread nuclear introgression of African rDNA and repetitive gene families that are genomically dispersed. More recently, the progenitors of *G. gossypioides* and *G. raimondii* came into sexual contact, resulting in cytoplasmic introgression into *G. gossypioides*, and possibly the transfer of less conspicuous nuclear genes.

the present-day geographic distributions of the parent species and their interfertility relationships with *G. gossypioides*. Our rationale for this hypothesis is detailed below.

### *ITS and Repetitive DNA from G. gossypioides Point to Widespread Nuclear Introgression from an African Cotton*

The phylogenetic resolution of *G. gossypioides* offered by nDNA contrasts the extraordinary results derived from ITS (Wendel et al. 1995) and from the distribution of repetitive elements (Zhao et al. 1998). In noting the anomalous phylogenetic placement and composition of the ITS sequence from *G. gossypioides*, these authors invoked hybridization with an African, A-genome taxon, and suggested that the ITS sequence of *G. gossypioides* represented a recombined chimera of A- and D-genome specific nucleotides. Re-analysis here with a more appropriate outgroup shows that the ITS of *G. gossypioides* is not chimeric (Table 5); instead, seven apparent synapomorphies unite *G. gossypioides* with A-, B-, and F-genome African cottons, and *no* apparent synapomorphies unite this species with D-genome cottons (Table 4).

In their evaluation of *G. gossypioides* ITS, Buckler et al. (1997) correctly predicted that the similarity to D-genome sequences was entirely due to plesiomorphic characters. However, these authors discounted introgression as the source of the anomalous ITS sequence, favoring instead a scenario whereby two divergent rDNA arrays exist in diploid cottons, one which is G/C rich (class I) and the other which is A/T rich (class II). According to this hypothesis, the affinity of *G. gossypioides* to other African cottons is best explained by biased sampling of paralogous ITS sequences during PCR amplification. Internal transcribed spacer sequences from *G. gossypioides*, A-, B-, and F-genome African cottons are relatively A/T rich, are thus characterized as class II, and as such constitute a clade of orthologous sequences that are sister to (paralogous) class I sequences isolated from all other cottons.

This ''rDNA paralogy'' hypothesis remains a possibility, especially since in situ hybridization has revealed three major and two minor 45S rDNA arrays in diploid cottons (Hanson et al. 1996). This ''rDNA paralogy'' hypothesis nevertheless appears contradicted by the most recent estimate of the cotton phylogeny, which is based on an extensive molecular data set (~ 18 kb of data; Cronn et al. 2002) and more appropriate rooting than previous studies (Wendel et al. 1995; Cronn et al. 1996; Small et al. 1998). If class II ITS sequences are indeed paralogous to class I sequences, then the A-, B-, and F-genome African cottons (all class II) should form a monophyletic sister group to all class I sequences, the latter of which would include all remaining species in the genus. Analysis of ITS using exemplars from across the cotton genus (Cronn et al. 2002) supports a monophyletic A-, B-, an F-genome lineage, but this clade is not sister to all remaining (class I) species. Instead, these African cottons are sister to Australian C-, G-, and K-genome groups, and this monophyletic Old World clade resolves sister to New World cottons (*cf.* Fig. 3; Cronn et al. 2002). This phylogenetic res-

olution is identical to the pattern exhibited by six mapped low-copy nuclear genes (*A1341*, *AdhA*, *CesA1b*, *FAD-2-1*, *G1121*, and *G1134*; Cronn et al. 2002). The near-perfect topological correspondence between ITS and low-copy markers (*G. gossypioides* is the sole exception) suggests that orthological relationships of ITS arrays from these species are not dramatically different from low-copy markers (where orthology has been established by linkage mapping).

Our hypothesis of African introgression would be difficult to support if it were based solely upon nuclear ITS sequences, especially given the increasing reports regarding the unpredictable phylogenetic behavior of rDNA (summarized in Àlvarez and Wendel 2003). In this context, the study by Zhao et al. (1998) on the repetitive DNA content of *G. gossypioides* and other diploid cottons provides an independent assessment of the ''Africanized'' composition of the *G. gossypioides* nuclear genome. Their survey of diploid cottons for 83 unique repetitive DNA families showed *G. gossypioides* to be unique among D-genome cottons, possessing about 400,000 copies of repeats from 20 distinct classes that are ''African genome specific'' (restricted to A-, B-, E-, and F- genomes) or ''African genome enriched.''

Because *G. gossypioides* occupies a basal phylogenetic position in the New World clade, it could be argued that the common progenitor of all cottons shared these elements, and that they have been independently retained in *G. gossypioides* and African cottons yet simultaneously purged from all remaining New World and Australian species. In situ hybridization of these repeats shows that ''African-specific/enriched'' repeats are distributed across all chromosomes in the *G. gossypioides* genome (Zhao et al. 1998). If the progenitor of modern cottons showed a similar genomic distribution of repetitive elements as *G. gossypioides*, the independent and genome-wide elimination of these 20 repetitive DNA families from the divergent New World and Australian cotton lineages (Fig. 1) is difficult to explain mechanistically. In contrast, genome-wide transfer of repetitive DNA from A-genome chromosomes to D-genome chromosomes in tetraploid cotton has already been demonstrated for replicative elements (Zhao et al. 1998; Hanson et al. 1998, 1999), and similar phenomena have been described for interspecific hybrids of diploid rice (Liu and Wendel 2000). If widespread mobilization of repetitive DNA is a common response to the ''genomic shock'' imposed by hybridization, then the presence of numerous ''African specific/enhanced'' elements in *G. gossypioides* provides tantalizing evidence of an ancient hybridization event with an (now extinct) African species.

If this hypothesis—namely, introgressive hybridization between the ancestor of modern *G. gossypioides* and an African species (outlined in Fig. 5)—accurately describes how *G. gossypioides* acquired these genomically dispersed ''African'' repetitive elements, we can only speculate on the mechanisms by which this introgression might occur, particularly because modern American and African diploids are intersterile and have chromosomes that differ in size by a factor of two (Endrizzi et al. 1985). We note, however, that theoretically improbable hybridization events are increasingly reported (Rieseberg 1997; Canter et al. 1999; Faure et al. 2002). It is possible that introgression was accompanied by massive gene conversion involving repetitive elements, and that introgressed low-copy nuclear elements may also exist which have not yet been sampled. Noteworthy in this regard is that *G. gossypioides* synthesizes terpenoids that are restricted to African cottons (Altman et al. 1990), suggesting that the genes responsible for this biosynthetic capacity may be the product of such an introgression event. Irrespective of the mechanism(s) responsible for repetitive element introgression, the net influence on genome size would be negligible, because *G. gossypioides* retains the small genome of its D-genome relatives (2 pg/2C) rather than the larger genome of African cottons (3.4–3.8 pg/2C; Endrizzi et al. 1985). Notwithstanding our ignorance regarding mechanism, it seems likely that interspecific genetic exchange in plants could be far more widespread than recognized, possibly entailing mechanisms that do not fit conventional models requiring chromosome pairing and balanced gametes for fertility.

If the ITS from *G. gossypioides* is indeed African in origin, it cannot be assigned to an extant cytogenetic group because it shares one apparent synapomorphy each with B-, and E-genome cottons (Seelanan et al. 1997; Table 4). By contrast, A-genome ITS sequences from tetraploid AD-genome *G. mustelinum* share seven synapomorphies with African A-, B-, and F-genome cottons, and *seven additional synapomorphies* with A-genome diploids (Wendel et al. 1995). This suggests that the putative source of rDNA introgression into *G. gossypioides*, although African, was not the A-genome progenitor of New World tetraploid cottons (contra Wendel et al. 1995), because the ITS sequence from that ancestral species was clearly ''A-genome-like'' at the time of A × D hybridization and polyploid formation. Instead, African introgression into *G. gossypioides* involved a taxon that was genetically equidistant from modern-day A-, B-, and F-genome cottons.

Finally, if our proposed hybridization scenario is correct, it is significant in that it points to a second, previously unrecognized African species that transiently colonized the New World. Long distance dispersal events of this kind—although unimaginably rare—are not without precedent in *Gossypium*, because the maternal genome donor of New World AD-genome tetraploid cotton originated in Africa (Wendel and Albert 1992; Cronn et al. 1996; Small and Wendel 1999). The A-genome of tetraploid cottons is nearly indistinguishable from extant A-genome diploid species (Brubaker et al. 1999; Cronn et al. 1999), and molecular clock estimates place the divergence of African A- and F-genomes at less than one million years before present (Cronn et al. 2002). This recent divergence indicates that trans-Atlantic dispersal of African cotton propagules to the New World occurred at least once, and it is a viable explanation for the existence of ''African'' repetitive DNA in *G. gossypioides*. As with the A-genome diploid progenitor of tetraploid cottons, this unknown African cotton would have become established in the New World, would have hybridized with a resident (*G. gossypioides*–like) cotton species, and subsequently have become extinct. Also like the African progenitor to tetraploid cottons, the sole evidence of this African species' existence in the New World appears to be hidden in nucleotide sequences of its hybrid descendent, *G. gossypioides*.

### *cpDNA from G. gossypioides Indicates a Recent Introgression and Cytoplasmic Transfer Event from an American Diploid Cotton*

Ancient introgression from an extinct African cotton may reconcile the conflicting message obtained from repetitive sequences (Zhao et al. 1998), ITS (Wendel et al. 1995) and nDNA gene trees from *G. gossypioides*, but it does little to resolve the apparent incongruence between nDNA and cpDNA data. Although it is possible to invoke lineage sorting of a cytoplasmically polymorphic ancestor to account for the apparent sister relationship between *G. gossypioides* and *G. raimondii* based on cpDNA, several observations suggest that such an explanation may be unlikely. First, the cladistic divergence between *G. gossypioides* and *G. raimondii* indicated by multiple nuclear genes would require the retention of a shared cytoplasmic polymorphism through all speciation events subsequent to the origin of the New World lineage (Figs. 2, 3). Even if this polymorphism were to exist at the time of the New World radiation, the small and scattered population sizes characteristic for most *Gossypium* species (Fryxell 1988), their predominant self-pollination (Wendel and Cronn 2003), and the small effective population size of haploid chloroplast loci relative to diploid nuclear loci would tend to drive such polymorphism rapidly to fixation. In light of the minimal cpDNA sequence divergence between cotton species of a common subsection (e.g., Wendel and Albert 1992), lineage sorting seems an unlikely source for generating the phylogenetic conflict evident in *G. gossypioides*.

If lineage sorting can be rejected as a potential explanation for the phylogenetic incongruence in the cpDNA data, then introgressive hybridization of a foreign cpDNA genome into *G. gossypioides* would be the most likely alternative (Fig. 5). Given the topologies recovered in cpDNA-based trees (Figs. 2, 3), the timing of this introgression event is implicated as cladistically recent, because sequence divergence between *G. gossypioides* and *G. raimondii* is exceedingly low (0.20%) across 7.3 kb of sequence. Whereas introgression from *G. raimondii* (Peru) into *G. gossypioides* (Oaxaca, Mexico) seems biogeographically implausible, *G. raimondii* likely derived from the core cotton group that presently inhabits western Mexico. For this reason, the historical ranges of their progenitors were certainly more proximate. Artificial hybridization experiments attest to the potential for natural hybridization between these species, because *G. raimondii* is the only species examined with which *G. gossypoides* will form fertile hybrids (Menzel and Brown 1954).

### *Prevalence of Hybridization and Introgression in Gossypium*

Although this report has focused on the striking cpDNA–rDNA–nDNA incongruence in *G. gossypioides*, cpDNA-nDNA incongruence is also evident in the lineage composed of the two closely related species *G. davidsonii* and *G. klotzschianum* (Fig. 2), which are presently restricted to Baja California and the Galapagos Islands, respectively. This species pair is resolved in cpDNA-based topologies (Fig. 2; Wendel and Albert 1992) as basal among D-genome cottons. This placement conflicts with morphological evidence (Fryxell 1971) and nuclear DNA sequence data (Fig. 2; Cronn et al. 1996; Seelanan et al. 1997; Small and Wendel 2000a; Liu et al. 2001), the latter unanimously supporting a close phylogenetic relationship between *G. davidsonii*/*G. klotzschianum*, and *G. raimondii*. We note that one interpretation of the conflicting cpDNA and nDNA gene trees from New World cottons is that *G. gossypioides* possesses what may be described as a ''*G. davidsonii*/*klotzschianum*-like'' plastome (i.e., similar to that found in *G. raimondii*), whereas *G. davidsonii*/*klotzschianum* possess a ''*G. gossypioides*-like'' plastome. These two entities are now geographically (Baja California to Oaxaca) and reproductively (Brown and Menzel 1952) isolated, but the suggestion emerges that there may have been a cytoplasmic ''swap'' between compatible progenitors in a region of sympatry outside of their present range. In this hypothesized hybrid zone, progenitors of the modern lineages acquired introgressed cytoplasms that ultimately achieved fixation in the descendant species *G. gossypioides* and *G. davidsonii*/*G. klotzschianum*. If this interpretation is correct, then as with other aspects of the history of *G. gossypioides*, the present implausibility of genetic exchange between geographically and genetically disparate lineages underscores the remarkable phenomenon of historical introgression between lineages for which sexual contact seems impossible.

If our interpretations of the *G. gossypioides* incongruence are correct, then interspecific hybridization has been more important in the diversification of *Gossypium* than previously recognized. In addition to the taxa discussed in this study, analyses based on twelve nuclear and four chloroplast genes (Cronn et al. 2002) revealed similar topological conflict in the resolution of the African B-genome lineage. Similar incongruence has also been described for the Mexican species *G. aridum* (Wendel and Albert 1992) and the Australian species *G. bickii* (Wendel et al. 1991; Seelanan et al. 1999; Liu et al. 2001) and *G. cunninghamii* (Seelanan et al. 1999). When combined with the five AD-genome allotetraploid species and the putatively introgressed species identified in this paper, 12 of 45 extant *Gossypium* species may have a recent reticulate ancestry. This high frequency is unexpected, particularly because reticulation has been considered unimportant in the evolution of diploid *Gossypium* (Fryxell 1971, 1979, 1992), most of which preferentially self-pollinate and commonly occur in small, scattered populations with no evident means of interspecific sexual contact. The geographic isolation of modern, apparently introgressant species also requires long-distance (even intercontinental) dispersal of one parent as a prerequisite to sexual contact. Moreover, the genetic and genomic incompatibilities that presently exist in the relevant species must have somehow been circumvented.

It is tempting to generalize, based on evidence from *Gossypium*, that hybridization and reticulation play an even greater role in plant speciation than presently thought, notwithstanding recognition of the importance of polyploidy (Masterson 1994; Wendel and Doyle 1998; Soltis and Soltis 1999) and the prevalence of hybrids in plants (Arnold 1997; Rieseberg 1997). Clearly, modern geographic and genetic barriers cannot be assumed to have been operative in the past, and thus ancient reticulation has occurred between lineages for which present distributions and reproductive biology suggest gene exchange is impossible. Additional examples of chimeric species (such as *G. gossypioides*) undoubtedly will

be revealed through comparative phylogenetic analyses in other taxa. An intriguing challenge in these cases will be to document not only the complex history of the lineages involved, but to elucidate the pattern of trait expression in such hybrids (Arnold 1997; Rieseberg 1997; Schwarzbach et al. 2001), the frequency of novel, transgressive phenotypic expression (Rieseberg 1997; Schwarzbach et al. 2001), and the adaptive significance of interspecific genetic exchange.

## Acknowledgments

The authors thank Corrinne Grover for technical assistance and Anna Gardner for the illustration shown in Figure 5. Keith Adams, Curt Brubaker, David Baum, Roberta-Mason Gamer and an anonymous reviewer provided constructive comments on the manuscript. This work was supported by grants from the Iowa State University Plant Science Initiative and the National Science Foundation.

## Literature Cited

Altman, D. W., R. D. Stipanovic, and A. A. Bell. 1990. Terpenoids in foliar pigment glands of the A, D, and AD genome cottons: introgression potential for resistance. J. Hered. 81:447–454.

Àlvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phyl. Evol. *In press*.

Arnold, M. L. 1997. Natural hybridization and evolution. Oxford Series in Ecology and Evolution, New York.

Barker, F. K., and F. Lutzoni. 2002. The utility of the incongruence length difference test. Syst. Biol. 51:625–637.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogeny reconstruction. Evolution 42:795–803.

Brown, M. S., and M. Y. Menzel. 1952. The cytology and crossing behavior of *Gossypium gossypioides*. Bull. Torr. Bot. Club 79: 110–125.

Brubaker, C. L., A. H. Paterson, and J. F. Wendel. 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42:184–203.

Buckler, E. S., IV, A. Ippolito, and T. P. Holtsford. 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. Genetics 145:821–832.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Canter, P. H., I. Pasakinskiene, R. N. Jones, and M. W. Humphreys. 1999. Chromosome substitutions and recombination in the amphiploid *Lolium perenne* × *Festuca pratensis* cv Prior ($2n = 4X = 28$). Theor. Appl. Genet. 98:809–814.

Cao, Y., J. Adachi, and M. Hasegawa. 1998. Comment on the quartet puzzling method for finding maximum-likelihood tree topologies. Mol. Biol. Evol. 15:87–89.

Cronn, R. C., and J. F. Wendel. 1998. Simple methods for isolating homoeologous loci from allopolyploid genomes. Genome 41: 756–762.

Cronn, R. C., X. Zhao, A. H. Paterson, and J. F. Wendel. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. J. Mol. Evol. 42:685–705.

Cronn, R. C., R. L. Small, and J. F. Wendel. 1999. Duplicated genes evolve independently after polyploid formation in cotton. Proc. Natl. Acad. Sci. USA 96:14406–14411.

Cronn, R. C., R. L. Small, T. Haselkorn, and J. F. Wendel. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Am. J. Bot. 89:707–725.

Dolphin, K., R. Belshaw, C. D. L. Orme, and D. L. J. Quicke. 2000. Noise and incongruence: interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17:401–406.

Endrizzi, J. E., E. L. Turcotte, and R. J. Koehl. 1985. Genetics, cytology, and evolution of *Gossypium*. Adv. Genet. 23:271–375.

Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics 10:315–320.

Faure, N., H. Serieys, A. Berville, E. Cazaux, and F. Kaan. 2002. Occurrence of partial hybrids in wide crosses between sunflower (*Helianthus annuus*) and perennial species *H. mollis* and *H. orgyalis*. Theor. Appl. Genet. 104:652–660.

Fryxell, P. A. 1971. Phenetic analysis and the phylogeny of the diploid species of *Gossypium* L. (Malvaceae). Evolution 25: 554–562.

———. 1979. The natural history of the cotton tribe. Texas A&M Press, College Station, TX.

———. 1988. Malvaceae of Mexico. Systematic Botany Monographs, vol. 25. American Society of Plant Taxonomists, Ann Arbor, MI.

———. 1992. A revised taxonomic interpretation of *Gossypium*. L. Rheedea 2:108–165.

Hanson, R. E., M. Nurul Islam-Faridi, E. A. Percival, C. F. Crane, Y. Ji, T. D. McKnight, D. M. Stelly, and H. James Price. 1996. Distribution of 5S and 18S–28S rDNA loci in tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105:55–61.

Hanson, R. E., X.-P. Zhao, M. N. Islam-Faridi, A. H. Paterson, M. S. Zwick, C. F. Crane, T. D. McKnight, D. M. Stelly, and H. J. Price. 1998. Evolution of interspersed repetitive elements in *Gossypium* (Malvaceae). Am. J. Bot. 85:1364–1368.

Hanson, R. E., M. N. Islam-Faridi, C. F. Crane, M. S. Zwick, D. G. Czeschin, J. F. Wendel, T. D. McKnight, H. J. Price, and D. M. Stelly. 1999. Ty1-*copia*-retrotransposon behavior in a polyploid cotton. Chromosome Res. 8:73–76.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Liu, B., and J. F. Wendel. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. Genome 43:874–880.

Liu, Q., C. L. Brubaker, A. G. Green, D. R. Marshall, P. J. Sharp, and S. P. Singh. 2001. Evolution of the FAD2-1 fatty acid desaturase 5′ UTR intron and the molecular systematics of *Gossypium* (Malvaceae). Am. J. Bot. 88:92–102.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science 264:421–424.

Menzel, M. Y., and M. S. Brown. 1954. Isolating mechanisms in hybrids of *Gossypium gossypioides*. Am. J. Bot. 42:49–57.

Osborn, T. C., J. C. Pires, J. A. Birchler, D. L. Auger, Z. J. Chen, H.-S. Lee, L. Comai, A. Madlung, R. W. Doerge, V. Colot, and R. A. Martienssen. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19:141–147.

Posada, D., and K. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Raymond, O., F. Piola, and C. Sanlaville-Boisson. 2002. Inferences of reticulation in outcrossing allopolyploid taxa: caveats, likelihood, and perspectives. Trends Ecol. Evol. 17:3–6.

Rieseberg, L. J. 1997. Hybrid origins of plant species. Annu. Rev. Ecol. Syst. 28:359–389.

Schwarzbach, A. E., L. A. Donovan, and L. H. Rieseberg. 2001. Transgressive character expression in a hybrid sunflower species. Am. J. Bot. 88:270–277.

Seelanan, T., A. Schnabel, and J. F. Wendel. 1997. Congruence and consensus in the cotton tribe (Malvaceae). Syst. Bot. 22:259–290.

Seelanan, T., C. L. Brubaker, J. McD. Stewart, L. A. Craven, and J. F. Wendel. 1999. Molecular systematics of Australian *Gossypium* section *Grandicalyx* (Malvaceae). Syst. Bot. 24:183–208.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114–1116.

Small, R. L., and J. F. Wendel. 1999. The mitochondrial genome of allotetraploid cotton (*Gossypium* L.). J. Hered. 90:251–253.

———. 2000a. Phylogeny, duplication, and intraspecific variation of *Adh* sequences in New World diploid cottons (*Gossypium*, Malvaceae). Mol. Phyl. Evol. 16:73–84.

———. 2000b. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155:1913–1926.

Small, R. L., J. A. Ryburn, R. C. Cronn, T. Seelanan, and J. F. Wendel. 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogenetic reconstruction in a recently diverged plant group. Am. J. Bot. 85:1301–1315.

Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. 14:348–352.

Strimmer, K., and A. von Haesler. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

Swofford, D. L. 1998. PAUP*, phylogenetic analysis using parsimony (*and other methods), Vers. 4.0. Sinauer, Sunderland, MA.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37:221–244.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

Wendel, J. F. 2000. Genome evolution in polyploids. Plant Mol. Biol. 42:225–249.

Wendel, J. F., and V. A. Albert. 1992. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17:115–143.

Wendel, J. F. and R. C. Cronn. 2003. Polyploidy and the evolutionary history of cotton. Adv. Agron. 78:139–186.

Wendel, J. F., and J. J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. Pp. 265–296 *in* P. S. Soltis, D. E. Soltis, and J. J. Doyle, eds. Molecular systematics of plants. II. DNA sequencing. Kluwer, Dordrecht, The Netherlands.

Wendel, J. F., and A. E. Percival. 1990. Molecular divergence in the Galapagos Islands–Baja California species pair, *Gossypium klotzschianum* and *G. davidsonii* (Malvaceae). Plant Syst. Evol. 171:99–115.

Wendel, J. F., J. McD. Stewart, and J. H. Rettig. 1991. Molecular evidence for homoploid reticulate evolution among Australian species of *Gossypium*. Evolution 45:694–711.

Wendel, J. F., A. Schnabel, and T. Seelanan. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. Mol. Phyl. Evol. 4: 298–313.

Wendel, J. F., R. C. Cronn, J. S. Johnston, and H. J. Price. 2002. Feast and famine in plant genomes. Genetica 115:37–47.

Zhao, X.-P., Y. Si, R. E. Hanson, C. F. Crane, H. J. Price, D. M. Stelly, J. F. Wendel, and A. H. Paterson. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Res. 8:479–492.

Corresponding Editor: D. Baum